AD-A038 944 WISCONSIN UNIV MADISON MATHEMATICS RESEARCH CENTER F/G 12/1

REGRESSION WITH GIVEN MARGINALS. (U)

JAN 77 R A VITALE DAAG29-75-C-0024

UNCLASSIFIED MRC-TSR-1710 NL

1 OF 1

AD A038944

END

DATE FILMED

5-77

(12)

MRC Technical Summary Report # 1710

REGRESSION WITH GIVEN MARGINALS

Richard A. Vitale

Mathematics Research Center
University of Wisconsin–Madison
610 Walnut Street
Madison, Wisconsin 53706

See 1473

January 1977

(Received December 21, 1976)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

REGRESSION WITH GIVEN MARGINALS

Richard A. Vitale

Technical Summary Report # 1710
January 1977

ABSTRACT

We consider the class of regression functions $\mathcal{M}(F, G) = \{m(x) = E[Y|X = x], (X, Y) \in \Pi(F, G)\}$ where $\Pi(F, G)$ denotes the set of random vectors with marginal distributions F and G. A characterization of $\mathcal{M}(F, G)$ is given together with a representation for the projection operator it induces in an appropriate Hilbert space. Applications are indicated.

AMS (MOS) Subject Classifications: Primary 62J05; Secondary 28A65, 46C10, 60G25

Key Words: **Regression, isotonic regression, convex minorant, rearrangement of a function, nonlinear prediction**

Work Unit Number 4 (Probability, Statistics, and Combinatorics)

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

REGRESSION WITH GIVEN MARGINALS

Richard A. Vitale

1. Introduction

Let $\Pi(F, G)$ denote the class of random vectors (X, Y) with marginal distributions F and G $(X \sim F,\ Y \sim G)$. We will consider the associated class of regression functions

$$\mathcal{M}(F, G) = \{m(x) = E[Y \mid X = x],\ (X, Y) \in \Pi(F, G)\} .$$

The motivation for looking at this class is similar in spirit to that of isotonic regression (from which we will in fact borrow a result): the extent to which auxiliary information be incorporated into the regression process. Knowledge of marginal distributions, in particular, is natural in certain types of problems. We may consider a census in which bivariate observations are collected, the marginal distributions are assumed given (as from a previous survey), and regression is desired. Alternatively, there is the problem of optimal, non-linear prediction in a time series $\{X_i\}$. If F is the equilibrium distribution of the X_i, then the optimal one-step predictor (squared error loss) is $E[X_{i+1} \mid X_i = x] \in \mathcal{M}(F, F)$ (see [3], [5], [6] for related discussions of this problem).

In section 2, we present a characterization of $\mathcal{M}(F, G)$ for a large class of F and G. The proof follows directly

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

from methods in [10]. Characterizations of the type indicated have been investigated from a variety of points of view and we refer the reader to [7], [9] for other discussions and references. It can be fairly stated that the common ancestor of all such approaches is the fertile theorem of Hardy, Littlewood and Pólya [4, p. 49] on the averaging properties of doubly stochastic matrices. In section 3, we investigate further the structure of $\mathcal{M}(F, G)$ by considering it as a convex subset of an appropriate Hilbert space and examining the induced projection operator. The discussion is motivated by a statistical estimation problem.

2. Characterization of $\mathcal{M}(F, G)$

In what follows we shall regard F and G as fixed and satisfying

(A1) F and G are each supported on all of R^1 and are invertible.

(A2) $EY^2 = \int_{-\infty}^{+\infty} y^2 G(dy) < \infty$.

The first assumption can be weakened considerably, but we present it to avoid side-issues. The second insures that $\mathcal{M}(F, G)$ is a subset of $L_2[(-\infty, +\infty); F]$, the Hilbert space of real-valued functions on R^1 square integrable with respect to the measure determined by F (this can be seen directly by noting $EY^2 = E_X E[Y^2|X] \geq E_X(E[Y|X])^2$).

Turning to the characterization of $\mathcal{M}(F, G)$, we note that if $m(x) = E[Y|X = x] \in \mathcal{M}(F, G)$, *then with the application* of marginal probability transformations $U = F(X)$, $V = G(Y)$, we have $m(x) = E[G^{-1}(V)|U = F(x)]$, where U and V are each uniformly distributed on $[0, 1]$. This is essentially the object of study of [10] and with only minor modifications, the methods employed there yield the following result.

<u>Theorem 1</u>. The following statements are equivalent.

(i) $m \in \mathcal{M}(F, G)$.

(ii) m lies in the closed convex hull $(L_2[(-\infty, +\infty); F])$ of functions of the form $G^{-1} \circ T \circ F$.

(iii) $\int_0^x m(F^{-1}(T(u)))du \geq \int_0^x G^{-1}(u)du$

for all $x \in [0, 1]$ (with equality at $x = 1$) and all $T \in \mathfrak{T}$.

Here $\mathfrak{J} = \{T : [0,1] \to [0,1]$ one-one, Borel-measurable, measure-preserving$\}$. We note that if $m \circ F^{-1}$ is non-decreasing, then the strongest inequality in (iii) occurs upon taking $T(u) = u$, i.e.,

$$\int_0^x m(F^{-1}(u))du \geq \int_0^x G^{-1}(u)du .$$

The equality condition in (iii) amounts to $\int_{-\infty}^{+\infty} m(x)F(dx) = \int_{-\infty}^{+\infty} yG(dy)$ or $Em(X) = EY$. Finally, for the projection problem it will be useful to note that the mapping $h \in L_2[(-\infty,+\infty);F] \to h \circ F^{-1} \in L_2[[0,1];\mu =$ Lebesgue measure$]$ induces an isomorphism between the two spaces. The image $\mathcal{M}_0$ of $\mathcal{M}(F,G)$ under the mapping can be described as follows.

<u>Corollary</u>. The following are equivalent.

(i) $m_0 \in \mathcal{M}_0$.

(ii) m_0 lies in the closed convex hull $(L_2[[0,1];\mu])$ of functions of the form $G^{-1} \circ T$.

(iii) $\int_0^x m_0(T(u))du \geq \int_0^x G^{-1}(u)du$

for all $x \in [0,1]$ (with equality at $x = 1$) and all $T \in \mathfrak{J}$.

<u>Proof</u>. Change of variables.

<u>Remark</u>. From (ii), it is evident that for each $T \in \mathfrak{J}$, $m_0 \in \mathcal{M}_0 \Longleftrightarrow m_0 \circ T \in \mathcal{M}_0$.

3. Projection

Under the assumption $(X, Y) \in \Pi(F, G)$, a natural criterion for judging an estimate $\hat{m}(x)$ of the unknown regression function $m(x)$ is the squared error loss

$$E[m(x) - \hat{m}(x)]^2 = \int_{-\infty}^{+\infty} [m(x) - \hat{m}(x)]^2 F(dx) .$$

It is evident that this loss can be reduced (or at least made no larger) by constructing a new estimate $\tilde{\hat{m}}(x)$ which is the projection of $\hat{m}$ onto the convex $\mathcal{M}(F, G)$. For this reason, it is of interest to investigate the projection operator associated with $\mathcal{M}(F, G)$ in $L_2[(-\infty, +\infty); F]$: that is, for $h \in L_2[(-\infty, +\infty); F]$, we seek the (unique) element $\tilde{h} \in \mathcal{M}(F, G)$ which yields

$$\int_{-\infty}^{+\infty} [h(x) - \tilde{h}(x)]^2 F(dx) = \inf_{m \in \mathcal{M}(F,G)} \int_{-\infty}^{+\infty} [h(x) - m(x)]^2 F(dx)$$

($\sim$ throughout will denote projection in the appropriate space). A feature of this projection is that if a constant is added to h, then $\tilde{h}$ remains the same: this can be seen by expanding

$$\int_{-\infty}^{+\infty} [h(x) + c - m(x)]^2 F(dx) = \int_{-\infty}^{+\infty} [h(x) - m(x)]^2 F(dx)$$

$$+ c^2 + 2c \int_{-\infty}^{+\infty} h(x) F(dx)$$

$$- 2c \int_{-\infty}^{+\infty} m(x) F(dx)$$

and noting that the first term alone depends on m since, as we have noted, $\int_{-\infty}^{+\infty} m(x)F(dx) \equiv \int_{-\infty}^{+\infty} yG(dy)$ for $m \in \mathcal{M}(F,G)$. This being the case, we shall have occasion to invoke the normalization

$$(A3) \quad \int_{-\infty}^{+\infty} h(x)F(dx) = \int_{-\infty}^{+\infty} yG(dy)$$

and, equivalently, for $\ell = h \circ F^{-1}$

$$(A3)' \quad \int_0^1 \ell(u)du = \int_0^1 G^{-1}(u)du \,.$$

We now investigate the projection operator, isolating the main aspects of the argument in two lemmas. Some notation will prove to be convenient: let $I(x) = \int_0^x G^{-1}(u)du$ and let capitalization generally indicate integration, e.g. $L(x) = \int_0^x \ell(u)du$. If $A(x) \in C[0,1]$, then denote by $A^*(x)$ the convex minorant of A (i.e. the greatest convex function less than or equal to A).

<u>Lemma</u>. Let $\ell \in L_2[[0,1];\mu]$ be non-decreasing (a.e.) and satisfy (A3)'. The projection $\tilde{\ell}$ of ℓ onto $\mathcal{M}_0$ satisfies

$$\tilde{L}(x) = \int_0^x \tilde{\ell}(u)du = L(x) - (L - I)^*(x) \,.$$

<u>Proof</u>. The proof will be given first for step functions and then extended.

(I) For a fixed integer $N \geq 1$, suppose that ℓ is of the form

$$\ell(u) = \sum_{j=0}^{N-1} \ell_j I_{[x_j, x_{j+1}]}(u), \quad x_j = \frac{j}{N}, \ \ell_j \leq \ell_{j+1} \,.$$

We argue first that it is enough to restrict attention to candidates for projection which are similarly non-decreasing step functions: given $n \in \mathcal{M}_0$, we apply the Cauchy-Schwarz inequality to get

$$\int_0^1 [\ell(u) - n(u)]^2 du = \sum_{j=0}^{N-1} \int_{x_j}^{x_{j+1}} [\ell_j - n(u)]^2 du \geq \sum_{j=0}^{N-1} \frac{1}{N} (\ell_j - n_j)^2$$

where $n_j = N \int_{x_j}^{x_{j+1}} n(u)du$. The lower bound is attained for $n(u)$ identically constant on sub-intervals. Moreover, it can further be reduced by rearranging the n_j to be non-decreasing ([4, theorem 378]). If $n_j^{(T)}$ are the rearranged values, then we have

$$\int_0^1 [\ell(u) - n(u)]^2 du \geq \int_0^1 [\ell(u) - n^{(T)}(u)]^2 du$$

where $n^{(T)}(u) = \sum_{j=0}^{N-1} n_j^{(T)} I_{[x_j, x_{j+1}]}(u)$. We now show that $n^{(T)}(u) \in \mathcal{M}_0$. Since $n^{(T)}(u)$ is non-decreasing (a.e.), by the remark after theorem 1, it is enough to show that $N^{(T)}(x) = \int_0^x n^{(T)}(u)du \geq I(x)$ with equality at $x = 1$. The latter condition follows from the normalization (A3)'. Since $I(x)$ is convex and $N^{(T)}(x)$ is piece-wise linear, it is enough to verify the inequality constraints at the nodes $\{x_j\}$. We have

$N^{(T)}(x_k) = \int_0^{x_k} n^{(T)}(u)du = \frac{1}{N} \sum_{j=0}^{k-1} n_j^{(T)}$, which is the integral of $n(u)$ over k of the sub-intervals. Equivalently, it is equal to $\int_0^{x_k} n(T(u))du$

for some T which appropriately permutes the sub-intervals. By (ii) of the corollary, this is bounded from below by $I(x_k)$.

We now have a discrete problem to solve:

$$\text{minimize} \quad \sum_{j=0}^{N-1}{}' (\ell_j - n_j)^2$$

subject to (a) the n_j are non-decreasing,

and (b) $\sum_{j=0}^{k-1}{}' n_j \geq I(x_{k-1})$, $k = 1, \ldots, N-1$ with equality at $k = N$.

Imposing only constraint (b), the problem is treated in [1, pp. 46-51] as a generalized isotonic regression. Letting L and $\tilde{L}$ denote the partial sum vectors of ℓ and the solution vector $\tilde{\ell}$ respectively and setting $I = (I(x_1), I(x_2), \ldots, I(x_N))$, we have

$$\tilde{L} = L - (L - I)^*$$

where $*$ here denotes the convex minorant of a vector. A straightforward argument shows that $\Delta_k^2(L - I)^* \leq \Delta_k^2(L - I)$ (Δ_k^2 denoting a second difference). Hence

$$\Delta_k^2 \tilde{L} = \Delta_k^2[L - (L - I)^*] = \Delta_k^2 L - \Delta_k^2(L - I)^* \geq \Delta_k^2 I \geq 0 .$$

It follows that $\tilde{L}$ is convex and that $\tilde{\ell}$ is non-decreasing. Thus (a) is satisfied automatically.

Translating the solution of the discrete problem into step function terms, we get $\tilde{L}(x) = L(x) - (L - I)^*(x)$.

(II) If $\ell(u)$ is not a step function, then for each $N \geq 1$, approximate $\ell(u)$ with

$$\ell_N(u) = \sum_{j=0}^{N-1} [N \int_{x_j}^{x_{j+1}} \ell(u)du] \, I_{[x_j, x_{j+1}]}(u) .$$

By (I), we have

$$\tilde{L}_N(x) = L_N(x) - (L_N - I)^*(x) . \tag{1}$$

Now as $N \to \infty$, $\ell_N \to \ell$ and $\widetilde{\ell_N} \to \tilde{\ell}$ in $L_2[[0,1];\mu]$. Since $[\int_0^x \ell_N(u)du]^2 \leq \int_0^x \ell_N^2(u)du \to \int_0^x \ell^2(u)du$, the dominated convergence theorem yields $L_N(x) \to L(x)$. Similarly, $\tilde{L}_N(x) \to \tilde{L}(x)$. Further, since $L_N \to L$ uniformly and $*$ operates continuously in the uniform norm, $(L_N - I)^* \to (L - I)^*$. Taking limits $(N \to \infty)$ in (1) yields the lemma.

If ℓ is not monotone, then some additional preparation is required to obtain its projection on $\mathcal{M}_0$. For $\ell \in L_2[[0,1];\mu]$, define $\ell_\uparrow \in L_2[[0,1];\mu]$ as the increasing rearrangement of ℓ. There exists a measure-preserving transformation $S_\ell : [0,1] \to [0,1]$, <u>not necessarily one-one</u>, such that $\ell = \ell_\uparrow \circ S_\ell$ ([8]).

<u>Lemma</u>. Let $\ell \in L_2[[0,1];\mu]$ and satisfy (A3)'. Then if ℓ and $\widetilde{\ell_\uparrow}$ are the projections of ℓ and $\ell_\uparrow$ respectively onto $\mathcal{M}_0$,

$$\tilde{\ell} = \widetilde{\ell_\uparrow} \circ S_\ell .$$

<u>Remark</u>. The construction for $\widetilde{\ell_\uparrow}$ has been given in the previous lemma.

<u>Proof</u>. If $\ell \in L_2[[0,1];\mu]$, then $\ell_\uparrow \in L_2[[0,1];\mu]$. Using a change of

variables, we have

$$\int_0^1 [\ell_\uparrow(u) - g(u)]^2 du = \int_0^1 [\ell(u) - (g \circ S_\ell)(u)]^2 du$$

and taking infima over $g \in \mathcal{M}_0$

$$\int_0^1 [\ell_\uparrow(u) - \ell_\uparrow(u)]^2 du = \inf_{g \in \mathcal{M}_0} \int_0^1 [\ell(u) - (g \circ S_\ell)(u)]^2 du$$

$$= \int_0^1 [\ell(u) - (\widetilde{\ell}_\uparrow \circ S_\ell)(u)]^2 du .$$

The lemma will follow if we can show

(i) $$\inf_{g \in \mathcal{M}_0} \int_0^1 [\ell(u) - (g \circ S_\ell)(u)]^2 du = \inf_{g \in \mathcal{M}_0} \int_0^1 [\ell(u) - g(u)]^2 du$$

and

(ii) $\widetilde{\ell}_\uparrow \circ S_\ell \in \mathcal{M}_0$.

Each is a consequence of the identity $\mathcal{M}_0 \circ S_\ell = \mathcal{M}_0$, that is, $g \circ S_\ell \in \mathcal{M}_0 \Longleftrightarrow g \in \mathcal{M}_0$. The point of interest is that S_ℓ may not be one-one. However, Brown [2, theorem 3] has shown that there exists a sequence $\{T_n\} \subseteq \mathfrak{J}$ such that $g \circ T_n \to g \circ S_\ell$. Accordingly, if $g \in \mathcal{M}_0$, then $g \circ T_n \in \mathcal{M}_0$ (see the remark after the corollary of section 1) and since $\mathcal{M}_0$ is closed $\lim_{n \to \infty} g \circ T_n = g \circ S_\ell \in \mathcal{M}_0$. Conversely, if $g \circ S_\ell \in \mathcal{M}_0$, then using an approximating sequence $\{T_n\}$

$$\|g \circ S_\ell - g \circ T_n\|_{L_2[[0,1];\mu]} = \|g \circ S_\ell \circ T_n^{-1} - g\|_{L_2[[0,1];\mu]} \to 0 .$$

Since $g \circ S_\ell \circ T_n^{-1}$ for each n and $\mathcal{M}_0$ is closed, we have $g \in \mathcal{M}_0$.

We can now state our main result.

Theorem 2. Let $h \in L_2[(-\infty, +\infty);F]$ and satisfy (A3). Let $(h \circ F^{-1})_\uparrow$ be the increasing rearrangement of $h \circ F^{-1}$ with $h \circ F^{-1} = (h \circ F^{-1})_\uparrow \circ S$. Then the projection $\tilde{h}$ of h onto $\mathcal{M}(F, G)$ is given by

$$\tilde{h} = \widetilde{(h \circ F^{-1})_\uparrow} \circ S \circ F$$

where $\widetilde{(h \circ F^{-1})_\uparrow}$ satisfies

$$\int_0^x \widetilde{(h \circ F^{-1})_\uparrow}(u)du = J_1(x) - J_2^*(x)$$

and $J_1(x) = \int_0^x (h \circ F^{-1})_\uparrow (u)du$, $J_2(x) = J_1(x) - \int_0^x G^{-1}(u)du$.

Proof. Together with the indicated isomorphism between $L_2[[0,1];\mu]$ and $L_2[(-\infty, +\infty);F]$, the statement combines the two lemmas.

4. Concluding Remarks

We have investigated the structure of $\mathcal{M}(F, G)$ through a characterization result and an examination of the induced projection operator. Despite the rather formidable description of the latter, computational versions have proved to be accessible. In particular, the operations $*$ and $\dagger$ together with the extraction of the measure-preserving transformation S are reasonably straightforward (a discussion of some relevant algorithms can be found in [1]).

As in isotonic regression, the fact that analytical resources are available to attack the problem investigated here suggests that other nonlinear regression problems may be amenable to similar treatment.

REFERENCES

[1] Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972). Statistical Inference Under Order Restrictions. Wiley, New York.

[2] Brown, J. R. (1966). Approximation theorems for Markov operators. Pacific J. Math. 16, 13-23.

[3] Grenander, U., McClure, D. E., and Vitale, R. A. (1969). Prediction. DAM and CCIS Report, Brown University.

[4] Hardy, G. H., Littlewood, J. E., and Pólya, G. (1967). Inequalities (second edition). Cambridge University Press.

[5] Jaglom, A. M. (1971). Examples of optimal nonlinear extrapolation of stationary random processes. Selected Transl. Math. Statist. and Prob. 9, 273-298.

[6] Masani, P., and Wiener, N. (1959). Non-linear prediction. In Probability and Statistics, U. Grenander ed., 190-212.

[7] Ryff, J. V. (1965). Orbits of L^1-functions under doubly stochastic transformations. Trans. Am. Math. Soc. 117, 92-100.

[8] Ryff, J. V. (1970). Measure preserving transformations and rearrangements. J. Math. Anal. Appl. 31, 449-458.

[9] Strassen, V. (1964). The existence of probability measures with given marginals. Ann. Math. Stat. 36, 423-439.

[10] Vitale, R. A., and Pipkin, A. C. (1976). Conditions on the regression function when both variables are uniformly distributed. Ann. Prob. 4, 869-873.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 1710	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER: 9 Technical	
4. TITLE (and Subtitle): 6 REGRESSION WITH GIVEN MARGINALS	5. ~~TYPE OF REPORT~~ & PERIOD COVERED: Summary Report - no specific reporting period
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): 10 Richard A. Vitale	8. CONTRACT OR GRANT NUMBER(s): 15 DAAG29-75-C-0024
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Mathematics Research Center, University of Wisconsin, 610 Walnut Street, Madison, Wisconsin 53706	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: U. S. Army Research Office, P.O. Box 12211, Research Triangle Park, North Carolina 27709	12. REPORT DATE: 11 Jan[redacted] 1977
	13. NUMBER OF PAGES: 13
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 12 16p.	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report) 14 MRC-TSR-1710

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Regression
isotonic regression
convex minorant
rearrangement of a function
nonlinear prediction

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We consider the class of regression functions $\mathcal{M}(F,G) = \{m(x) = E[Y|X = x], (X,Y) \in \Pi(F,G)\}$ where $\Pi(F,G)$ denotes the set of random vectors with marginal distributions F and G. A characterization of $\mathcal{M}(F,G)$ is given together with a representation for the projection operator it induces in an appropriate Hilbert space. Applications are indicated.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 221 200

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

KB